

Etimologia não é morfologia: produtividade e composicionalidade na formação e processamento dos “compostos morfológicos” do português

João Veloso & Pedro Tiago Martins

Faculdade de Letras da Universidade do Porto

Centro de Linguística da Universidade do Porto¹

Abstract

Words that combine different etyma do not correspond necessarily to morphologically complex words. We argue that the main property for classifying a word as morphologically complex is its compositional processing. Portuguese lexicon admits a set of words that combine different Latin or Greek etyma. These are commonly said to form a special subset within the lexicon of Portuguese and to be the result of a special process of compounding (“morphological compounding”), defined on the basis of etymology only. Since etymological information is not part of the speakers’ I-language (the linguist’s explicandum), these words should be described as morphologically simple words.

Keywords: morphology, etymology, productivity, morphological compounds,

Palavras-chave: morfologia, etimologia, produtividade, compostos morfológicos

1. Introdução: questão central e pressupostos teóricos

Neste trabalho, ocupar-nos-emos da seguinte questão específica: até que ponto poderemos aceitar como pertinentes para a descrição linguística *sincrónica* do português, ou de qualquer outra língua, certos aspetos da forma e da estrutura das palavras atribuíveis **unicamente** à etimologia e à evolução histórica da língua? Concretizando a pergunta de forma mais detalhada: será aceitável basear certas categorizações linguísticas ou individualizar determinados processos tidos como gramaticais com base, exclusivamente, em critérios ou aspetos históricos que não têm representação, no estágio atual da língua, no conhecimento linguístico implícito dos falantes?

Para a discussão desta questão, iremos debruçar-nos sobre um conjunto lexical do português supostamente individualizado² quanto ao processo de formação das palavras que o integram: os “*compostos morfológicos*”.

Partiremos, no nosso estudo, dos seguintes pressupostos de base:

Textos Seleccionados, XXVI Encontro da Associação Portuguesa de Linguística, Lisboa, APL, 2011, pp. 558-573

¹ Unidade U0022/2003 da Fundação para a Ciência e a Tecnologia (Programa FEDER/POCTI), que financiou parcialmente esta investigação.

² Villalva (2000; 2008) e Mateus *et al.*, (2003) contam-se entre as obras que identificam a “composição morfológica” como um processo de formação de palavras individualizado em português e as palavras dele resultantes como um subconjunto delimitado dentro do léxico da língua.

1) O objeto de descrição/explicação do linguista é a **gramática implícita dos falantes** (a entidade que Chomsky (1986) designa como a **língua-I**)³.

2) Dessa gramática implícita **não consta a informação explícita acerca da etimologia ou da história da língua**, visto que essa informação:

2.1) **não é comum a todos os falantes da língua**;

2.2) **não é acessível à intuição linguística dos falantes**;

2.3) **é adquirida em contexto formal explícito**, em resultado de uma experiência cultural não universal que é a escolarização;

2.4) e **não é inteiramente produtiva** (isto é, não cria novas palavras na língua nem interfere com o processamento de palavras, ou pelo menos não o faz de forma tão sistemática, tão frequente e tão partilhada pelos falantes da língua como o faz o conhecimento linguístico *implícito*).

2. Criatividade, produtividade e composicionalidade na gramática implícita dos falantes: formação e processamento de palavras

Um dos argumentos mais comumente apresentados como evidência da gramática implícita reside na *criatividade linguística*: de forma completamente independente da aprendizagem formal, da consciência explícita e do armazenamento de itens individualizados em resultado da experiência linguística prévia, a língua-I, finita, gera, processa e avalia um número infinito de produtos verbais inteiramente novos e potencialmente desconhecidos dos falantes (cf., entre outros: Chomsky, 1959: 576-577 e ss.; 1965: 6 e ss.; 1986: 48, 67 e ss.; Higginbotham, 1990: 246; Pinker, 1990: 207-208; 1994: 84 e ss.; Duarte, 2001). É com base nesta evidência que podemos considerar que todos os processos gramaticais representados cognitivamente – isto é, implantados na língua-I dos sujeitos – deverão apresentar-se como **produtivos** (ou seja, capazes de gerar e processar um número potencialmente infinito de produtos linguísticos a partir de um número finito de recursos) para poderem ser aceites como parte integrante do sistema computacional que é a gramática interiorizada dos falantes descrito pelos linguistas. No caso concreto da formação de palavras, que aqui nos interessa particularmente, temos de admitir que uma parte importante da informação morfológica incluída na competência linguística dos falantes é a que contempla o conhecimento do conjunto finito de regras gerais que operam de forma previsível sobre o repertório de unidades morfológicas e lexicais a partir das quais se constroem as palavras da língua. É com base nestes pressupostos que diversos autores postulam a existência do subconjunto de regras gramaticais responsáveis, especificamente, pela formação de palavras: as *Regras de Formação de Palavras* – RFPs – propostas desde trabalhos como Halle (1973) e Aronoff (1976), p. ex.

³ Ao longo do presente texto, usaremos designações como *conhecimento (implícito) da língua*, *conhecimento linguístico (implícito)*, *língua-I*, *competência (linguística)* e *gramática implícita* para fazermos referência a este objeto cognitivo que cumpre aos linguistas descrever e explicar, de acordo com os pressupostos teóricos assumidos neste trabalho.

Tais regras, como qualquer outra componente da gramática implícita dos falantes, atuarão a nível das três manifestações externas da língua-I atrás mencionadas: *produção*, *processamento* e *avaliação* de todos os produtos verbais produzidos e percebidos pelos falantes.

Um mecanismo importante incorporado no conhecimento linguístico implícito dos falantes e ativado no processamento semântico de certas palavras é a *composicionalidade*: em palavras dotadas de radicais morfologicamente complexos, a computação do significado de tais radicais baseia-se no Princípio de Composicionalidade, que permite calcular (e torna predizível) o significado partilhado por toda a palavra através da identificação, em separado, do significado de cada um dos seus constituintes básicos. Sendo um princípio com origem na semântica (atribuído a Frege, 1892), a morfologia recupera-o precisamente para explicar o processamento de unidades morfologicamente complexas, como se torna patente na seguinte citação:

“A distinção entre palavras simples e complexas estabelece uma divisão fundamental no domínio das estruturas morfológicas, a que se deve associar uma outra, que traça um contraste entre palavras complexas com uma estrutura composicional e palavras complexas lexicalizadas. **Composicionalidade** é um princípio estabelecido para a semântica, segundo o qual o significado de um todo é uma função do significado das partes. Em morfologia, composicionalidade é um princípio de análise das palavras complexas, que identifica aquelas cujas propriedades são integralmente determinadas pela sua estrutura e pelas propriedades dos seus constituintes. Por outras palavras, a composicionalidade é uma propriedade das palavras cujas forma e interpretação são funções directas da forma e da interpretação dos seus constituintes.”

(Villalva 2008: 29)

As palavras formadas por derivação afixal, bem como os compostos morfossintáticos, exemplificam muito claramente a aplicação do Princípio de Composicionalidade ao processamento semântico. Com efeito, é admissível – com base, por exemplo, em afirmações como a de Villalva (2008: 29) acima transcrita – que a computação do significado dos radicais das palavras de (1) (palavras derivadas por afixação) e (2) (compostos morfossintáticos) assenta na combinação do significado de cada um dos *morfemas*⁴ encontrados no interior de todos os radicais pertencentes às

⁴ Usamos, neste passo, o termo morfema (cunhado por Baudouin de Courtenay na segunda metade do século XIX – cf., p. ex., Aronoff & Volpe, 2006) numa aceção próxima da definição tradicional que encontramos, p. ex., em Bloomfield (1933) – “a linguistic form which bears no partial phonetic-semantic resemblance to any other form, is a simple form or morpheme.” (Bloomfield, 1933: 161) –, isto é, como uma unidade segmentável e contínua (para o autor citado, processos morfológicos manifestados através de variação fonética, p. ex., não são considerados morfêmic).

Esta noção sofreu redefinições profundas posteriormente. Ao longo do resto do século XX, foram feitas tentativas de conciliação entre a categorização reducionista de Bloomfield (1933) e a importância gramatical de processos e unidades linguísticas não segmentais, nomeadamente por Harris (1942, 1951), Nida (1949) e Hockett (1968). O trabalho de Aronoff (1976) marca um ponto de viragem, mudando o foco da morfologia para o significado gramatical que tanto os morfemas de Bloomfield (1933) como as unidades linguísticas não contínuas partilham,

palavras em causa, conforme a análise morfológica aqui apresentada (no caso dos compostos morfossintáticos de (2), optámos pela separação unicamente de cada uma das palavras morfossintáticas encontradas no interior de cada qual).

(1) – Palavras do português passíveis de processamento composicional: palavras derivadas (por afixação)

[[[in][competent]]RadComplx[e]]Adjetivo

[[[re][começ]]RadComplx[a][r]]Verbo

[[[im][possível]]RadComplx]Adjetivo

[[[re][organiz]]RadComplx[a][ção]]Nome

(2) – Palavras do português passíveis de processamento composicional: compostos morfossintáticos

[[tenente]Nome[coronel]Nome]Nome

[[diretor]Nome[geral]Adj]Nome

Uma evidência importante que nos permite aceitar a hipótese do processamento composicional das palavras com radicais morfológicamente complexos encontra-se na suposição⁵ de que qualquer falante do português seria capaz de atribuir significado aos produtos verbais apresentados em (3) (sujeitos aqui a uma análise interna muito sumária, que se limita a separar constituintes como tema e afixos derivacionais, basicamente). Estes resultam da aplicação das mesmas RFPs responsáveis pela geração das palavras

colocando ambos os conceitos sob a alçada do termo *morfema*. A partir de então, altamente influenciados pelas inovações de Aronoff (1976), vários autores – incluindo o próprio – estudam o problema de maneira mais aprofundada, tentando chegar aos fundamentos daquilo que se pode entender por morfema. Alguns estudos importantes neste período são os de Matthews (1991), Anderson (1992), Halle & Marantz (1993) e Aronoff (1994). Anderson (1992) questiona a importância de um conceito de morfema material, e explora aspetos como a existência de estrutura sem significado, acabando por chegar à conclusão de que o morfema bloomfieldiano é irrelevante sincronicamente, dando preferência às Regras de Formação de Palavras de Aronoff (1976) como explicação morfológica (Anderson, 1992: 48). Halle & Marantz (1993) propuseram uma nova teoria, a Morfologia Distribuída, segundo a qual os morfemas são vistos como elementos terminais de árvores sintáticas, independentemente de terem ou não conteúdo fonológico e da sua especificidade. Assim, segundo os autores, os morfemas podem ser vistos como complexos de características morfológicas (Halle & Marantz, 1993: 114).

Aronoff (1994) introduziu um novo termo, *morfoma*, para se referir a unidades linguísticas completamente formais e vazias de significado. Elementos que se inserem nesta categoria são, por exemplo, os paradigmas flexionais partilhados por verbos não relacionados semanticamente.

⁵ Não dispomos, à data da redação deste texto, de estudos de campo que possam mostrar melhor de que forma se dá, efetivamente, o reconhecimento e o processamento das palavras consideradas neste estudo, nomeadamente no que diz respeito à ativação (ou não) do Princípio de Composicionalidade. Por essa razão, assumiremos em diversos momentos um conjunto de *suposições* ou *conjeturas* acerca desse aspeto, baseando-nos na nossa observação empírica e na nossa intuição de falantes nativos com conhecimento explícito das questões linguísticas tratadas, deixando para um estudo futuro uma exploração empírica desta vertente do problema (que incluirá desejavelmente uma observação controlada, em diversos grupos de sujeitos sociolinguisticamente delimitados, das várias manifestações desse processamento).

de (1) e (2). Diferem apenas num ponto: constituem exemplos de *sobregeração* (cf., p. ex., Spencer, 1991: 90), não sendo por isso, neste momento, palavras dicionarizadas do português (pelo que não se encontram armazenadas no léxico mental dos falantes). Como tal, o hipotético reconhecimento do seu significado só se torna explicável, precisamente, se aceitarmos que o Princípio de Composicionalidade tem algum tipo de representação na gramática implícita dos falantes do português.

(3) – Não palavras processáveis por falantes do português com base no Princípio de Composicionalidade

[[acaba]_{TV}[çã]_{SufxDeriv}Nome(NãoDicionarizado)
 [[re]_{PrefixDeriv}[amassar]_{Infinitivo}Verbo(NãoDicionarizado)
 [[faca]_{Nome}[agulha]_{Nome}Nome(NãoDicionarizado)

Para que o Princípio de Composicionalidade atue no processamento dos produtos verbais de (3), torna-se necessário, entre outros fatores, que a cada um dos respetivos constituintes (de acordo com uma análise estrutural como a aí apresentada) corresponda, na língua-I dos falantes do português, um significado (não composicional), posteriormente computado de modo composicional no processamento das unidades morfológicamente complexas em que tais constituintes ocorrem. O significado dessas unidades morfológicamente complexas torna-se, então, *predizível* a partir da sua forma, ou seja, a partir do alinhamento e combinação dos morfemas que as integram (num certo sentido, o processamento composicional das palavras assemelha-se ao processamento do significado frásico, igualmente composicional e igualmente aplicável a um número potencialmente infinito de produtos linguísticos originais).

Pelo contrário, o significado de palavras com radicais morfológicamente simples é completamente *impredizível* a partir da sua forma: o processamento semântico de palavras como “janela”, “parede” e “sala”, p. ex. (v. as análises propostas em (4), em que assumimos, com base em propostas tradicionais como as de Camara (1970; 1971), a vogal final como *vogal temática*⁶), baseia-se exclusivamente na ativação da informação idiossincrática correspondente a cada um dos seus radicais e como tal **armazenada** de forma holística no léxico mental dos falantes.

(4) – Palavras com radicais morfológicamente não complexos

[[janel]_{Rad}[a]_{VT}Nome
 [[pared]_{Rad}[e]_{VT}Nome
 [[sal]_{Rad}[a]_{VT}Nome

⁶ A assunção de uma vogal temática nominal nos termos aqui referidos será retomada posteriormente, neste mesmo trabalho, noutras análises morfológicas.

Devido às características que apresentámos – e tendo em atenção a sua efetividade no tocante ao processamento de palavras potenciais mas não dicionarizadas, conforme pretendemos ilustrar com os exemplos de (3) –, o Princípio de Composicionalidade contribui, pelo menos em parte, para a explicação de numerosos casos de inovação lexical na língua, facilitando a entrada de palavras como as de (5) no léxico português. Trata-se de neologismos relativamente recentes que combinam palavras ou unidades morfológicas pré-existentes e disponíveis no léxico partilhado por um número supostamente substancial de falantes. Podemos, por esta razão, afirmar que o Princípio de Composicionalidade é produtivo no português contemporâneo (como o será, em princípio, em todas as línguas cujas gramáticas admitam processos derivacionais e composicionais semelhantes aos do português).

(5) – Neologismos do português sujeitos a processamento composicional

[[[precar]_{Raiz}[iz]_{SufxDeriv}[a]_{VT}]_{Tema(NãoAtestadoEmVerbo)}[ção]_{SufxDeriv}]_{Nome}
 [[governo]_{Nome}[sombra]_{Nome}]_{Nome}
 [[contra]_{Prep}[informação]_{Nome}]_{Nome}

3. Produtividade e composicionalidade nos “compostos morfológicos”

Como afirmámos na introdução a este estudo, o nosso principal objetivo consiste na discussão da adequação de certas categorizações gramaticais com base unicamente em critérios de natureza estritamente histórica ou etimológica. Centraremos essa discussão no caso particular das palavras a que certas descrições linguísticas se referem como “*compostos morfológicos*”.

3.1. “Compostos morfológicos”: descrição

O léxico do português dispõe de um conjunto limitado de palavras, pertencentes quase exclusivamente ao vocabulário erudito, que combinam étimos diretamente importados, na sua maior parte, do latim clássico e do grego antigo, muitos deles recorrentes em mais do que uma palavra da língua. As gramáticas tradicionais e escolares apresentam frequentemente listas de tais étimos, com indicação do seu significado, designando por “compostos eruditos” as palavras que resultam da sua combinação. Um exemplo desta abordagem descritiva é encontrado na gramática de Cunha & Cintra (1984).

Referimo-nos, portanto, a palavras como as de (6) (onde a análise interna apresentada se limita a marcar uma fronteira entre étimos e, em alguns casos, a identificar uma vogal de ligação, não sendo, portanto, uma análise morfológica síncrona).

(6) – Exemplos de “compostos morfológicos” em português

[[biblio][teca]]	[[fratr][i][cida]]
[[biblio][filia]]	[[filo][sofia]]
[[fil][antropia]]	[[teo][sofia]]
[[eno][teca]]	[[teo][logia]]
[[rat][i][cida]]	[[fono][logia]]
[[hom][i][cida]]	[[morfo][logia]]
[[antropo][logia]]	[[cronó][metro]]

Propostas de descrição morfológica da língua como Villalva (2000) consideram que os étimos encontrados no interior destas palavras continuam a gozar do estatuto de *radicais*⁷ no português contemporâneo. De acordo com esta perspetiva, os “*compostos morfológicos*” correspondem a “estruturas resultantes de um processo de concatenação de radicais simples [...] ou complexos [...], autonomamente existentes na língua, ou não, por intermédio de uma vogal de ligação [...]” (Villalva 2000: 353)⁸.

3.2 . “Compostos morfológicos”: processamento

Nesta secção, analisaremos a questão específica da possibilidade de os “compostos morfológicos” serem ou não processados de acordo com o Princípio de Composicionalidade. A principal interrogação que formulamos é a seguinte: serão os falantes do português contemporâneo capazes de atribuir significados em separado aos étimos resultantes da análise apresentada em (6), tornando *previsível* o significado de cada um desses pretensos compostos a partir da sua análise interna (ou seja, a partir de um processamento de tipo composicional como o defendido para as palavras de (1) e (2)), ou, pelo contrário, serão as palavras constantes dessa lista sujeitas a um processamento holístico que atribui globalmente a cada uma um significado idiossincrático e *imprevisível a partir da sua forma*, de acordo com o seu

⁷ A categorização dos étimos combinados nestas palavras enquanto radicais e a importância deste facto para a defesa da individualização da “composição morfológica” enquanto processo de formação de palavras em português são enfatizadas, de forma muito explícita, na seguinte afirmação: “[u]ma das principais características da composição morfológica consiste no facto de se tratar de um processo de *concatenação de radicais*. Esta afirmação [...] assenta na constatação de que estas unidades se podem combinar livremente entre si, ocorrendo, em alguns casos, quer como o primeiro [...], quer como o último [...] constituinte da estrutura composta [...]” (Villalva, 2000: 355; itálico nosso).

⁸ Nas propostas descritivas de Villalva (2000) e Mateus *et al.*, (2003), na categoria dos “compostos morfológicos” têm cabimento três tipos especiais de palavras do português: (i) combinações de étimos clássicos em que cada étimo conserva alguma individualidade fonológica (cada um deles conserva acento próprio), como sucede em “geotecnia” ou “antropofagia”; (ii) combinações de étimos clássicos em que se dá uma completa fusão dos dois (ou mais) étimos numa só palavra fonológica, que atua como uma entrada única para as regras de atribuição de acento e redução do vocalismo átono, como sucede em “geografia” e “antropologia”; (iii) combinações de radicais disponíveis na língua como palavras dotadas de classe lexical, como sucede em “lusó-descendente”, “subsídio-dependente” ou “sintático-semântico”, por exemplo.

Neste trabalho, ocupar-nos-emos exclusivamente dos “compostos” contemplados pelo caso (ii): palavras morfossintáticas que combinam étimos importados de outras línguas, inexistentes no português enquanto palavras distintas e sujeitos a uma “fusão” completa numa só palavra morfossintática que serve de entrada única para as regras de atribuição de acento, esbatendo-se assim a fronteira entre os constituintes morfológicos respetivos. Pretendemos ocupar-nos do estatuto das palavras resultantes dos outros dois processos composicionais referidos num trabalho futuro.

armazenamento no léxico mental de cada falante (como sucede com as palavras não morfologicamente complexas de (4))?

Partiremos, nesta análise, da suposição de que muitos falantes do português terão armazenadas no seu léxico mental palavras como “agricultor”, “península” e “democracia”, p. ex. Contudo, aceitamos também a suposição de que apenas uma parcela dos falantes da língua será capaz de atribuir significado (não composicional) a segmentos infralexicais encontrados nessas palavras tais como “agr-”, “pen-”, “-ínsul-”, “demo-” e “-craci(a)”⁹. Isto é: poucos falantes armazenarão e processarão estas palavras de acordo com a análise apresentada em (6); falantes sem um conhecimento explícito da história da língua – isto é, a maioria dos falantes – não se mostrarão capazes de individualizações etimológicas como as encontradas em (6).

Um primeiro ponto que nos parece pertinente pôr em destaque nesta discussão é o seguinte: a descrição destas palavras como palavras com radicais morfologicamente complexos – radicais que combinam constituintes mais pequenos semanticamente individualizados – só faz sentido num conjunto minoritário de falantes¹⁰: nos que acederam a essa informação de forma explícita, em resultado da aprendizagem formal de aspetos estruturais e históricos da gramática da língua. O conhecimento etimológico é, em si mesmo, *intuível* e não faz parte da faculdade da linguagem que torna possível a língua-I, nos termos de Chomsky (1986) ou Raposo (1992), por exemplo (o português falado hoje é uma língua diferente do português medieval ou quinhentista, e diferente do latim ou do grego, e a informação respeitante a essas línguas não pode constar de forma alguma da gramática interna intuível de um falante do português contemporâneo).

Na definição do conceito de língua-I defendida em fontes clássicas da gramática generativa como Chomsky (1986), a forma como *todos* os falantes interiorizam as propriedades centrais da sua gramática implícita assenta na referida faculdade da linguagem e traduz-se num conjunto de generalizações intuíveis acerca da organização estrutural do sistema linguístico. Por consequência, a parte do conhecimento linguístico explícito eventualmente associada à formação (histórica) destas palavras apresenta duas características muito particulares e interligadas, que podemos de alguma forma

⁹ O estatuto morfológico preciso destes segmentos infralexicais é discutível. Na nota 4, fizemos referência a várias aceções de morfema em diversas correntes e épocas da história da morfologia. Mesmo aceitando que a definição tradicional do morfema como a menor unidade linguística dotada de significado apresenta limitações e que, em algumas propostas morfológicas (como em algumas que referimos na nota 4), são admissíveis morfemas destituídos de forma fonológica ou conteúdo semântico, a identificação de morfemas com base exclusivamente na etimologia das palavras (“morfemas etimológicos”) parece-nos, à luz dos argumentos desenvolvidos neste estudo, irrelevante para a descrição da gramática implícita dos falantes.

¹⁰ No subconjunto de falantes com conhecimento explícito deste aspeto da história do português, incluem-se, supostamente, os linguistas, o que nos poderia levar a ver propostas de análise que dão estas palavras como “complexas” como, até certo ponto, um efeito de “viés metodológico” (em que a perspetiva particular do observador o leva a divisar particularidades não presentes no objeto descrito). Para uma discussão sobre o lugar da introspeção enquanto metodologia em linguística (cognitiva), cf., por exemplo, González-Márquez *et al.*, (eds., 2007), bem como, num plano mais geral, as observações encontradas em Pinker (2008: 34).

Por outro lado, a questão coloca-se também, em nosso entender, porque estamos a trabalhar numa língua cuja história e cuja etimologia são conhecidas. Se estivéssemos a estudar uma língua de etimologia desconhecida (o basco ou uma língua ameríndia, por exemplo), a morfologia descritiva não se ocuparia de hipóteses como a possível composicionalidade dos radicais destas palavras nem chegaria à formulação de propostas de categorização gramatical como a que aqui se encontra em discussão.

considerar como “marginais” para a constituição da competência linguística dos falantes: trata-se de um conhecimento não partilhado por todos os falantes da língua e resulta de uma experiência cultural explícita e formal. Nestas duas características, esta parcela “marginal” do conhecimento da língua distingue-se claramente do formato “básico” da língua-I tal como ele tem sido entendido ao longo do texto, isto é, como um “estado da mente” (Chomsky, 1986: *passim*; 1995: 14) largamente invariável de sujeito para sujeito e altamente independente da experiência cultural dos indivíduos (Chomsky, 1986: 44 e ss., 62-63, 67 e ss.; 1988: 3-4 e ss., 15 e ss., 27, 34; 1999: 43 e ss.; Raposo, 1992: 35-36)¹¹.

Como tal, é lícito supor que estas palavras, no léxico mental de grande parte dos falantes da língua, nem conterão radicais morfológicamente complexos, nem serão, por isso, processadas de forma composicional¹². A sua composicionalidade – que, é certo, existiu em dado momento histórico e numa língua diferente do português contemporâneo¹³ – ter-se-á perdido por ação de um processo de **lexicalização**, entendido aqui nos termos em que é definido na seguinte citação de Villalva (2008), onde claramente se associa à perda de composicionalidade:

“Quanto à **lexicalização**, trata-se de um processo de perda da composicionalidade, que actua de forma aleatória e imprevisível, sempre que pelo menos um dos constituintes morfológicos sofre alterações semânticas ou formais *ou é desconhecido para os falantes*. A lexicalização pode, pois, afectar a interpretação da palavra, a sua forma, as suas propriedades gramaticais ou uma conjugação destes factores.”

(Villalva 2008: 30; *itálico nosso*)

Registe-se que este processo de lexicalização/perda de composicionalidade é um processo muito frequente na história do português. Muitas palavras que hoje são descritas como morfológicamente simples (isto é, dotadas de radicais morfológicamente simples) resultam da combinação no seu interior de dois (ou mesmo mais) étimos latinos distintos. Reunimos alguns exemplos em (7) (a análise dos radicais é aí meramente etimológica).

¹¹ Nesta distinção entre conhecimento linguístico “básico” e “marginal”, tentamos recuperar, de alguma forma, a distinção clássica entre “gramática nuclear” e “periferia” defendida por Chomsky (1981: 126-127) (e problematizada pelo mesmo autor em Chomsky, 1982: 110).

¹² É interessante sublinhar neste momento que na explicação da composicionalidade encontrada em Villalva (2008: 29) (vd. citação no ponto 2 deste texto), a autora admite a existência de palavras complexas composicionais e não composicionais (as lexicalizadas). Trata-se, na verdade, da distinção que pretendemos refutar neste trabalho, defendendo uma identificação entre composicionalidade e complexidade (já que a “complexidade” das palavras lexicalizadas só existe historicamente e, portanto, fora da língua-I dos falantes).

¹³ Não pretendemos, neste trabalho, desvirtuar o estudo da história do português ou de qualquer outra língua, que consideramos uma área de estudo fértil e pertinente e que fornece aos linguistas informações relevantes para a compreensão/explicação de certos aspetos verificados na língua contemporânea. O ponto central da nossa proposta situa-se ao nível da avaliação das intuições linguísticas dos falantes no processamento das palavras da língua, relativamente ao qual, de acordo com o ponto de vista desenvolvido neste estudo, a informação etimológica (não integrante da competência linguística) não desempenhará qualquer função.

(7) – Palavras descritas como dotadas de radicais morfologicamente simples em português que resultam da combinação de mais do que um étimo latino

seduzir: [[[se] Étimo1 [duz] Étimo2] Rad [i] VT [r] MMT_Inf] Verbo
percutir: [[[per] Étimo1 [cut] Étimo2] Rad [i] VT [r] MMT_Inf] Verbo
admitir: [[[ad]] Étimo1 [mit] Étimo2] Rad [i] MMT_Inf [r] Inf] Verbo
computador: [[[com] Étimo1 [puta] Étimo2] Rad [dor] Sufx_Deriv] Nome
república: [[[re] Étimo1 [públic] Étimo2] Rad [a] VT] Nome
benefício: [[[bene] Étimo1 [fíci] Étimo2] Rad [o] VT] Nome
florilégio: [[[flor] Étimo1 [i] VL [légi] Étimo2] Rad [o] VT] Nome
jurisconsulto: [[[juris] Étimo1 [consult] Étimo2] Rad [o] VT] Nome
cônjuge: [[[côn] Étimo1 [jug] Étimo2] Rad [e] VT] Nome

Levando ao extremo os argumentos que suportam a classificação das palavras de (6) como morfologicamente complexas, com base no facto de os constituintes morfológicos combinados no seu interior terem correspondido inquestionavelmente, no latim ou no grego, a radicais distintos, teríamos de considerar as palavras de (7) igualmente na categoria das palavras morfologicamente complexas, sujeitas por conseguinte a processamento composicional, e, consequentemente, candidatas à classificação como palavras morfologicamente “complexas”.

De um ponto de vista histórico, não restam quaisquer dúvidas de que cada uma destas palavras combina pelo menos dois étimos (conforme a análise etimológica dos seus radicais apresentada em (7)). Porém, e conforme vimos anteriormente, a sua composicionalidade perdeu-se totalmente em português (por *lexicalização*) e os falantes, na sua maioria, não se mostram conscientes nem da sua divisibilidade nos respetivos “constituintes etimológicos”, nem da possibilidade de atribuir significado a cada um deles separadamente. Sublinhamos, finalmente, que a combinação etimológica patente nestas palavras não corresponde a um processo inteiramente produtivo em português, já que muitas destas palavras entram já formadas em português a partir do próprio latim e porque em caso algum os morfemas do latim (ou de qualquer outra língua) podem ser um objeto sobre o qual operem as RFPs do português¹⁴.

Assim, é de admitir que estas palavras se distinguirão, ao nível do processamento a que são sujeitas pelos falantes do português, das palavras e não palavras de (1), (2) e (3), representadas/representáveis de forma composicional no léxico mental de todos os falantes da língua. Com efeito, o acesso intuitivo à computação composicional do significado dessas palavras e não palavras, verificado em *todos* os falantes e *independentemente da experiência cultural destes*, contrasta com a necessidade de *ativação do conhecimento explícito, formalmente adquirido*, para o processamento

¹⁴ Reconhecemos que, também neste aspeto, os “compostos morfológicos” cujo estatuto gramatical aqui se discute apresentarão uma diferença importante relativamente a estas palavras, já que boa parte deles (embora não a totalidade) pode ter sido formada em português.

composicional das palavras de (6) (e (7), eventualmente), **nos falantes que as processarem composicionalmente (e apenas nestes).**

Em suma, podemos conjecturar que, para um falante nativo do português, a decomposição de “antropologia” em [antrop(o)] e [logi(a)] (ou seja, decomposições coincidentes ou próximas das apresentadas em (6)) não é possível senão num subconjunto minoritário de falantes e mediante a ativação consciente do conhecimento explícito adquirido no âmbito da experiência escolar dos indivíduos¹⁵. Essa informação não é acessível às intuições linguísticas “genuínas” de *todos* os falantes. Para os falantes que não possuam conhecimento etimológico explícito, “antropologia” será porventura uma palavra tão composta (e tão composicional) como “parede” ou “janela”: todas elas detêm um significado atribuído globalmente a toda a palavra, não sujeito a qualquer tipo de análise morfológica interna.

Atendendo a todos os argumentos que expusemos relativamente à representação mental dos processos gramaticais, a grande divisão que nos parece justificada será entre palavras representadas e processadas de forma composicional e aquelas que não o são. Relacionando esta divisão com divisões tradicionais ou correntes nos estudos descritivos do português, deveríamos assim agrupar na categoria das palavras “composicionais” as derivadas por afixação e os compostos morfossintáticos e na das palavras “não composicionais” todas as restantes (inclusive aquelas que, nos termos referidos neste artigo, são frequentemente classificadas como “compostos morfológicos”). Relativamente a estas últimas, atendendo ao que dissemos acerca do seu processamento não composicional pela maior parte dos falantes e ao lugar que o conhecimento etimológico ocupa na gramática implícita, defendemos explicitamente que uma parte significativa destas palavras se encontre *lexicalizada* em português, como sucede com palavras como as de (7).

Uma outra dimensão do problema sobre que temos de refletir é o da *produtividade* da “composição morfológica”.

Tendo em mente (i) a limitação quantitativa do léxico formado pelos “compostos morfológicos”, (ii) o carácter discutível da sua representação composicional na maioria dos falantes do português, (iii) o facto de a informação etimológica só estar disponível em resultado de uma experiência cultural particular e, finalmente, (iv) que muitas destas palavras entram já formadas no léxico português (como empréstimos lexicalizados, diretamente importados do latim ou do grego ou através de outras línguas), podemos considerar a “composição morfológica” um processo escassamente produtivo na

¹⁵ De certa forma, podemos comparar o conhecimento etimológico ao conhecimento formalmente adquirido que diz respeito a objetos não linguísticos. Comparamo-lo aqui ao entendimento da matéria enquanto combinações de átomos, p. ex., que também está acessível apenas a um grupo restrito de pessoas que adquirem esse conhecimento por via da escolarização formal. Trata-se de um tipo de conhecimento enciclopédico adquirido por via da escolarização – e não através do mero contacto “intuitivo” com a realidade. Com base em todos os argumentos expostos e discutidos até ao momento, consideramos definitivamente que o conhecimento da etimologia das palavras é uma parcela desse conhecimento enciclopédico (formalmente adquirido, cumulativo e não produtivo) e não da competência linguística propriamente dita (intuitivamente adquirida, intuitivamente acessível e produtiva).

formação de palavras em português (na medida em que é diretamente responsável pela formação de uma quantidade muito diminuta de entradas lexicais).

No entanto, não podemos ignorar completamente que, num subconjunto porventura minoritário de falantes, a “composição morfológica” pode com efeito corresponder a um processo minimamente produtivo: o processo que permite o processamento de não palavras¹⁶ como “reamassar” e “acabação” (exemplificadas em (3)) seria o mesmo a permitir, nesse subconjunto minoritário de falantes, o processamento (de tipo composicional) de não palavras ou neologismos potenciais como os encontrados em (8) (mais uma vez, a análise dos exemplos é meramente etimológica e ignora bastantes aspetos da estrutura morfológica das palavras).

(8) – Palavras potenciais do português formadas por “composição morfológica” e passíveis de processamento composicional em falantes com conhecimento explícito da história da língua

[[sapato]][logia]]

[[filo]][grafia]]

[[fono]][metria]]

[[helio]][técnica]]

[[hipo]][fobia]]

Em conjugação com este facto, meramente hipotético, não podemos ignorar o conjunto, ainda que restrito, de alguns neologismos, como os indicados em (9), que demonstram que a “composição morfológica” retém alguma produtividade em português (novamente, os exemplos são sujeitos a uma análise etimológica sumária).

(9) – Neologismos do português formados por “composição morfológica” (“concatenação de radicais”) e passíveis de processamento composicional em falantes com conhecimento explícito da história da língua

[[skat]][ó]][dromo]]

[[samb]][ó]][dromo]]

[[queim]][ó]][dromo]]

[[bêdê]][teca]]

[[biblio]][bus]]

[[gadget]][ó]][mano]]

Admitimos que os indícios, ainda que limitados, de produtividade da “composição morfológica” que reconhecemos a propósito dos exemplos de (8) e (9) nos termos que

¹⁶ Usamos, no texto, a designação “não palavras” para nos referirmos a “palavras potenciais”, isto é, construções morfológica e fonotaticamente bem formadas mas não dicionarizadas na língua.

acabamos de expor podem vir a relativizar a nossa recusa de ver nestas palavras um processo diferenciado na gramática implícita dos falantes. A explicação que oferecemos para esta aparente contradição retoma um ponto já anteriormente focado: olhando às particularidades dos casos reunidos em (8) e (9), nomeadamente à obrigatoriedade de uma modelação cultural forte para que exista, num subconjunto de falantes apenas, um processamento composicional destas palavras, teremos sempre de conceber essa produtividade como um processo relativamente “marginal” – circunscrito, na gramática da língua, à sua **periferia** (e não ao seu **núcleo**), recorrendo de novo à distinção proposta por Chomsky (1981)¹⁷.

4. Considerações finais e desenvolvimentos futuros

Na primeira parte deste texto, afirmámos que as unidades e processos dotados de representação na língua-I dos falantes constituem o verdadeiro objeto de estudo em linguística e que só as unidades e processos relativamente aos quais disponhamos de evidência forte de tal representação constituem *explicanda* da investigação em linguística.

Assim, e com base nos argumentos expostos ao longo das secções 2 e 3 do estudo, defendemos:

1) que os “compostos morfológicos” não serão processados pela gramática implícita da maioria dos falantes de forma composicional. Por outras palavras, podemos dizer que o seu processamento não difere do das palavras com radicais simples (ao contrário do que sucede com os derivados afixais e com os compostos morfossintáticos, processados composicionalmente);

2) que estas palavras, na sua maior parte e na gramática implícita da maior parte dos falantes do português, foram já sujeitas a um processo de lexicalização – patente, precisamente, no processamento não composicional referido no parágrafo anterior e defendido ao longo de todo este texto;

3) que a “composição morfológica”, ainda que conservando alguns vestígios de produtividade (patentes, designadamente, a nível da formação de alguns neologismos associados a vocabulários específicos), não se pode considerar em português contemporâneo um processo verdadeiramente produtivo; a sua escassa produtividade regista-se somente num conjunto restrito de falantes da língua e requer um tipo de ativação consciente e culturalmente modelada que difere da ativação de outros processos gramaticais “regulares” na língua.

Como tal, a individualização dos “compostos morfológicos” numa classe à parte dentro do léxico do português com base no argumento de que estas palavras resultam da aplicação de um processo gramatical regular não encontra, a nosso ver, correspondência com a organização da gramática implícita dos falantes da língua no seu formato “nuclear” (vd. discussão sobre este aspeto em particular na secção 3.2 do texto). Não excluimos que, do ponto de vista da informação de natureza *enciclopédica* (isto é,

¹⁷ Vd. novamente nota 11.

“extralinguística”), o conhecimento da origem e do significado primitivo, na língua de origem, de cada um dos étimos destas palavras seja interessante para, a um dado nível de descrição explícita da língua, podermos aperceber-nos, nomeadamente, de algumas particularidades semânticas destas palavras. Todavia, este aspeto, porventura importante para a explicitação de algumas características da língua em certos patamares de observação externa da mesma, não é cognitivamente representado no conhecimento linguístico implícito, não sendo, por isso, atuante no processamento linguístico nem na estruturação da gramática implícita dos falantes. Consequentemente, reiteramos que a postulação de um processo gramatical regular responsável pela formação destas palavras e a concomitante categorização das mesmas numa classe à parte dentro do léxico do português parecem-nos carecer de fundamento à luz dos pressupostos teóricos enunciados no ponto 1 deste texto e centrais numa teoria generativa da faculdade da linguagem humana.

As reservas manifestadas neste estudo acerca da classificação dos “compostos morfológicos” num grupo à parte com base em argumentos estritamente associados apenas ao âmbito estrito da etimologia são extensíveis, em nossa opinião e pelas mesmas razões analisadas no trabalho, a um outro conjunto particular de palavras: os *pseudotruncamentos*, de que nos ocuparemos num estudo posterior. De acordo com autores como Araújo (2002) e Villalva (2008), entre outros, cabem nesta categoria as formas abreviadas que resultam da interrupção de um “composto morfológico” num ponto coincidente com a fronteira entre os dois étimos, como sucede nos exemplos de (10), retirados de Villalva (2008).

(10) – Exemplos de “pseudotruncamento” em português (exemplos de Villalva 2008: 59)

<i>foto</i>	fotografia
<i>metro</i>	metropolitano
<i>micro</i> _[-fem]	microfone
<i>micro</i> _[+fem]	micro-radiografia
<i>moto</i>	motocicleta
<i>porno</i>	pornográfico/a
<i>quilo</i>	quilograma
<i>zoo</i>	zoológico

Segundo os autores citados, esta estratégia seria, por um lado, produtiva, e, por outro, seria ativada por um suposto conhecimento etimológico dos falantes, combinado com a análise morfológica das palavras, como se depreende da seguinte explicação: “[o pseudotruncamento] parece reconhecer uma estrutura de composição morfológica, preservando o primeiro radical (que é geralmente um radical neoclássico) e a vogal de ligação” (Villalva 2008: 59).

Com base em todos os argumentos a que recorremos para propor que os “compostos morfológicos” resultam, na gramática implícita dos falantes, das mesmas RFPs responsáveis pela formação das palavras com radicais simples, já que dificilmente o conhecimento da etimologia se pode aceitar como parte da língua-I dos falantes, diremos, neste momento, que, em relação aos pseudotruncamentos, será útil equacionar a interferência de outras variáveis patentes na gramática implícita, nomeadamente as ligadas à estruturação prosódica e rítmica da fala. Como dissemos, porém, a questão específica dos pseudotruncamentos será objeto de um outro estudo, guiado pelas mesmas questões de partida e fundamentação teórica do presente.

Escolhemos para título deste texto uma afirmação que retoma a interrogação que Rosa (2009) escolhe para título de um artigo em que se debruça sobre o estatuto morfológico de processos como a abreviação em geral, excluindo-os do campo de análise da morfologia precisamente por eles envolverem mecanismos não inteiramente produtivos e discutivelmente presentes na gramática implícita dos falantes. Com base em todos os argumentos de fundo que percorremos ao longo deste estudo, reafirmamos, como principal conclusão final do presente trabalho, que, por muito relevante que a etimologia possa ser para a descrição/explicação de propriedades linguísticas, ela não cabe, de facto, no âmbito da morfologia enquanto componente da língua-I dos sujeitos. É só neste sentido que deve ser lida a afirmação “*etimologia não é morfologia*”.

Referências

- Anderson Stephen R. (1992) *A-morphous morphology*. Cambridge MA: Cambridge University Press.
- Araújo, Gabriel (2002) Truncamento e Reduplicação no Português Brasileiro. *Revista de Estudos da Linguagem, Belo Horizonte* 12, pp. 61-90.
- Aronoff, Mark (1976) *Word formation in generative grammar*. Cambridge MA: The MIT Press.
- Aronoff, Mark (1994) *Morphology by itself*. Cambridge MA: MIT Press.
- Aronoff, Mark & Mark Volpe (2006) Morpheme. In Keith Brown (org.) *The Encyclopedia of Language and Linguistics*. Oxford: Elsevier, pp. 274-276.
- Bloomfield, Leonard (1933) *Language*. New York: Henry Holt and Co.
- Camara Jr., Joaquim Mattoso (1970) *Estrutura da Língua Portuguesa*. Petrópolis RJ: Vozes.
- Camara Jr., Joaquim Mattoso (1971) *Problemas de Linguística Descritiva*. Petrópolis RJ: Vozes.
- Chomsky, Noam (1959) A Review of B. F. Skinner's *Verbal Behavior*. *Language* 35 (1), pp. 26-58. Reproduzido em J. A. Fodor & J. J. Katz (eds., 1964) *The Structure of Language*. Englewood-Cliffs NJ: Prentice-Hall, pp. 547-578.
- Chomsky, Noam (1965) *Aspects of the Theory of Syntax*. Cambridge MA: The MIT Press.
- Chomsky, Noam (1981) Markedness and Core Grammar. In Adriana Belletti, Luciana Brandi & Luigi Rizzi (eds.) *Theory of Markedness in Generative Grammar. Proceedings of the 1979 GLOW Conference*. Pisa: Scuola Normale Superiore di Pisa, pp. 123-146.
- Chomsky, Noam (1982) *Noam Chomsky on the generative enterprise: A discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht: Foris.
- Chomsky, Noam (1986) *Knowledge of language: its nature, origins and use*. New York: Praeger. Trad. Port.: *O conhecimento da língua: sua natureza, origem e uso*. Lisboa: Caminho, 1992.
- Chomsky, Noam (1988) *Language and Problems of Knowledge. The Managua Lectures*. Cambridge MA: The MIT Press.
- Chomsky, Noam (1995) *The Minimalist Program*. Cambridge MA: The MIT Press.
- Chomsky, Noam (1999) On the Nature, Use, and Acquisition of Language. In William C. Ritchie & Tej K. Bhatia (eds.) *Handbook of Child Language Acquisition*. San Diego CA: Academic Press, pp. 33-54.

- Cunha, Celso & L. F. Lindley Cintra (1984) *Nova gramática do português contemporâneo*. Lisboa: Sá da Costa. 5ª ed.
- Duarte, Inês (2001) Uso da Língua e Criatividade. In Fernanda Irene Fonseca, Isabel Margarida Duarte & Olívia Figueiredo (orgs.) *A Linguística na Formação do Professor de Português*. Porto: Centro de Linguística da Universidade do Porto, pp. 107-123.
- Frege, Gottlob (1892) Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* C, pp. 25-50.
- González-Márquez, Mónica, Irene Mittelberg, Seana Coulson & Michael J. Spivey (eds.) (2007) *Methods in Cognitive Linguistics*. Amsterdam: John Benjamins.
- Halle, Morris (1973) Prolegomena to a theory of word formation. *Linguistic Inquiry* 4 (1), pp. 3-16.
- Halle, Morris & Alec Marantz (1993) Distributed morphology and the pieces of inflection. In Kenneth Hale & Samuel Jay Keyser (eds.) *The view from building 20*. Cambridge MA: The MIT Press, pp. 111-176.
- Harris, Zellig (1942) Morpheme alternants in linguistic analysis. *Language* 18, pp. 169–180.
- Harris, Zellig (1951) *Methods in structural linguistics*. Chicago IL: The University of Chicago Press.
- Higginbotham, James (1990) Philosophical Issues in the Study of Language. In Daniel N. Osherson & Howard Lasnik (eds.) *An Invitation to Cognitive Science. Volume 1: Language*. Cambridge MA: The MIT Press/Bradford, pp. 243-257.
- Hockett, Charles (1968) *The state of the art*. The Hague: Mouton.
- Mateus, Maria Helena Mira *et al.*, (2003) *Gramática da Língua Portuguesa*. Lisboa: Caminho.
- Matthews, Peter H. (1991) *Morphology*. Cambridge: Cambridge University Press.
- Nida, Eugene (1949) *Morphology: the descriptive analysis of words*. Ann Arbor MI: University of Michigan Press.
- Pinker, Steven (1990) Language Acquisition. In Daniel N. Osherson & Howard Lasnik (eds.) *An Invitation to Cognitive Science. Volume 1: Language*. Cambridge MA: The MIT Press/Bradford, pp. 199-241.
- Pinker, Steven (1994) *The Language Instinct*. London: Penguin.
- Pinker, Steven (2008) *The Stuff of Thought*. London: Penguin.
- Raposo, Eduardo Paiva (1992) *Teoria da gramática. A faculdade da linguagem*. Lisboa: Caminho.
- Rosa, Maria Carlota (2009) É Morfologia? *Linguística – Revista de Estudos Linguísticos da Universidade do Porto* 4, pp. 45-58.
- Spencer, Andrew (1991) *Morphological Theory. An Introduction to Word Structure in Generative Grammar*. Oxford: Blackwell.
- Villalva, Alina (2000). *Estruturas morfológicas: Unidades e hierarquias nas palavras do português*. Lisboa: Fundação Calouste Gulbenkian.
- Villalva, Alina (2008) *Morfologia do Português*. Lisboa: Universidade Aberta.